

Conference paper

USGS Community for Data Integration (CDI): a community approach to managing science data

Leslie Hsu, U.S. Geological Survey

Summary

Data policies at the U.S. Geological Survey cover a broad range of geology, hydrology, and biology data types, and thus implementation of these open data policies is approached from many different viewpoints. The Community for Data Integration (CDI) helps to implement these policies with grassroots, member-driven activities, where USGS staff from varied scientific disciplines and geographically diverse regions collaborate together to solve data challenges. Through working groups, funded projects, meetings, trainings, and partnerships, CDI leads the development and adoption of collaborative tools and best practices that support the management, curation, and dissemination of the broad array of USGS data. External partnerships with other data organizations help to leverage products and exchange knowledge for the benefit of USGS data management. The CDI brings scientists, data managers, and IT professionals together, and the member-driven agenda allows for a dynamic and diverse environment that creates products and hosts activities that best serve the USGS data management needs.

Grassroots community activities to meet official policies

The U.S. Geological Survey maintains comprehensive policies and procedures for ensuring the quality and integrity of its science, including those related to the planning, collection, description, review, publication and preservation of data (U.S. Geological Survey, 2015a). Although policies regarding scientific data come from the leadership, implementation is often determined by practitioners at the grassroots level. The Community for Data Integration (CDI, <http://www.usgs.gov/cdi>) brings together expertise from representatives across the U.S. Geological Survey who are involved in research, data management, and information technology. By fostering an environment for collaboration and sharing, CDI allows its voluntary members to raise shared challenges and reach solutions collaboratively, thus improving the efficiency and effectiveness of their daily work in managing and manipulating scientific data.

The purpose of the CDI is to advance understanding of Earth systems through appropriate tools and techniques, provide a forum for people doing work with data integration to come together to share ideas and learn new skills and techniques, and grow overall USGS capabilities with data and information by increasing visibility of the work of many people throughout the USGS and the CDI community.

To achieve these goals, the CDI operates monthly forums, annual workshops, working groups, and funded projects (e.g., Chang et al., 2015; Langseth et al., 2015). The monthly

forums provide an open dialogue to share and learn about data integration efforts or to present problems that invite the community to offer solutions, advice, and support. An annual workshop facilitates the exchange of ideas, sharing of activities, presentations of current projects, and networking among members. Working groups evolve from common interests and are focused on efforts to address data management and technical challenges including the development of standards and tools, and improving data interoperability. Current CDI working groups include Citizen Science, CDI Communications, Connected Devices, Data Management, Earth-Science Themes, Semantic Web, and Technology Stack. An annual formal request for proposals (RFP) process funds projects that produce tangible products related to data management tools and practices, collaboration tools, and data integration and delivery. CDI's activities, which are collaborations between research scientists, data managers, and IT professionals, often directly address challenges related to implementation of new data policies at USGS, such as products that help with data management, workflow documentation, and data publication and preservation. In some cases, CDI products are adopted by the USGS for guiding future policy recommendations, such as the USGS Science Data Lifecycle Model (e.g., Faundeen et al., 2013), a framework for managing data.

Community connections help science data management and scientific research

While networking between CDI members helps advance information exchange for scientific data management, the CDI also looks outward to the many data-related organizations on the landscape and pursues opportunities to leverage products and community wisdom. A few examples are the Federation of Earth Science Information Partners (ESIP), the Research Data Alliance (RDA), DataOne, and EarthCube. Like CDI, each of these organizations hosts activities and working groups to highlight and address shared challenges in the community. Connections between these overlapping communities accelerate each group's progress beyond what they would achieve independently. Examples of community connections include jointly hosted workshops and hackathons, evaluation and use of working group products, and individual's cross-participation in multiple organizations. Through these connections, knowledge exchange between the communities mutually benefits all groups, and in particular, helps the USGS staff access resources and learn the skills needed to meet the USGS data management policies.

Competing Interests

The author declares that they have no competing interests.

References

Chang, M.Y., Carlino, Jennifer, Barnes, Christopher, Blodgett, D.L., Bock, Andy, Everette, A.L., Fernet, G.L., Flint, L.E., Gordon, Janice, Govoni, D.L., Hay, L.E., Henkel, H.S., Hines, M.K., Holl, S.L., Homer, Collin, Hutchison, V.B., Ignizio, D.A., Kern, Tim, Lightsom, F.L., Markstrom, S.L., O'Donnell, Michael, Schei, J.L., Schmid, L.A., Schoephoester, K.M., Schweitzer, P.N., Skagen, S.K., Sullivan, D.J., Talbert, Colin, and Warren, M.P., 2015, Community for data integration 2013 annual report: U.S. Geological Survey Open-File Report 2015–1005, 36 p., <http://dx.doi.org/10.3133/ofr20151005>.

USGS Community for Data Integration (CDI): a community approach to managing science data

Faundeen, J.L., Burley, T.E., Carlino, J.A., Govoni, D.L., Henkel, H.S., Holl, S.L., Hutchison, V.B., Martín, Elizabeth, Montgomery, E.T., Ladino, C.C., Tessler, Steven, and Zolly, L.S., 2013, The United States Geological Survey Science Data Lifecycle Model: U.S. Geological Survey Open-File Report 2013–1265, 4 p., <http://dx.doi.org/10.3133/ofr20131265>.

Langseth, M.L., Chang, M.Y., Carlino, Jennifer, Birch, D.D., Bradley, Joshua, Bristol, R.S., Conzelmann, Craig, Diehl, R.H., Earle, Paul, Ellison, L.E., Everette, A.L., Fuller, Pam, Gordon, J.M., Govoni, D.L., Guy, M.R., Henkel, H.S., Hutchison, V.B., Kern, Tim, Lightsom, F.L., Long, J.W., Longhenry, Ryan, Preston, T.M., Smith, Stan, Viger, R.J., Wesenberg, Katherine, and Wood, Eric, 2015, Community for Data Integration 2014 annual report: U.S. Geological Survey Open-File Report 2015–1184, 40 p., <http://dx.doi.org/10.3133/ofr20151184>.

U.S. Geological Survey, 2015a, USGS Fundamental Science Practices: U.S. Geological Survey, accessed May 23, 2016, at <https://www.usgs.gov/fsp/>.